International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)
ISSN (P): 2249–6831; ISSN (E): 2249–7943
Vol. 11, Issue 2, Dec 2021, 19–28

ADVANCE TECHNIQUE FOR EARLY DETECTION OF BREAST CANCER USING TEXTUAL ANALYSIS FROM DIGITAL MAMMOGRAM

SHAWNI DUTTA[1], ASMITA RAY[2], VISHAL GOYAL[2] & SAMIR KUMAR BANDYOPADHYAY[2]

[1]*The Bhawanipur Education Society College*

[2] *Professor, GLA University, Mathura-Delhi Road Mathura, Chaumuhan, Uttar Pradesh*

ABSTRACT

The field of image processing gaining importance is not only for its rapid and continuous progress but also for accurate and advanced analysis. Mammography is the most popular imaging technique for the detection of breast cancer Anatomical structure of a lesion is obtained properly compared to other imaging modalities like CT(Computed Tomography), MRI (Magnetic Resonance Imaging), PET (Positron-emission tomography). In this work, an algorithm has been developed for the detection of breast cancer. The proposed method has consisted of three steps: preprocessing, segmentation and feature extraction. After segmentation of cancerous region, it is characterized with statistical features using first-order histogram and Gray Level Co-occurrence Matrix (GLCM)). Based on these two types of feature extraction methods, normal and cancerous mammograms have been diagnosed.

KEYWORDS: Mammogram, CT, MRI, PET, Statistical features, GLCM

Received: May 26, 2021; **Accepted:** Jun 04, 2021; **Published:** Jul 05, 2021; **Paper Id.:** IJCSEITRDEC20213

Original Article

INTRODUCTION

Breast cancer is the most common and significant reason for death among women globally [1]. It is the second leading cause of death among women. Mammography is the most popular and well-known promising technique used by radiologists frequently for breast cancer detection. Detection and diagnosis of breast cancer at an early stage is the optimal solution for preventing tumor progression and successful medical intervention as the cause of this kind of disease remain almost unknown. Prevention of unwanted growth of malignant cells plays an important role to save the life of the patients. Different kinds of breast abnormalities are there like masses, micro-calcifications, speculated lesions and architectural distortion. Benign and malignant types can be present in all these kinds of abnormalities. Uncontrolled division of cell tissue is the main cause of breast cancer. The formation of large lump of tissues from abnormal cells consequently becomes tumor [2-4].

Micro-calcifications and masses in mammogram are considered as earliest signs of breast cancer. Micro-calcifications are the tiny flecks of calcium present inside the soft breast tissue [5]. Mass is the lump or tumor. A mass with a round, smooth and circumscribed boundary is called as benign mass and malignant tumor can be identified by speculated, rough and blurry boundary. Early detection increases the chance of recovery.

In 2012 nearly 1.7 million new cases of breast cancer has been diagnosed. In 2017, it is estimated that about 30% of newly diagnosed cancers in women will be breast cancers. Eighty-one percent of breast cancers are diagnosed among women ages 50 years and older, and 89% of breast cancer deaths occur in this age group [6].

LITERATURE REVIEW

R. Guzman Caberrera et al. [7] have been proposed an algorithm which is composed of background and objects texture segmentation and extraction. Local entropy has been used to separate suspicious region containing the masses from background parenchyma. Pavel kral et al. [8] proposed a method for breast cancer detection at an early stage, using LBP features. The background image and breast image are separated using Otsu method based on uniform LBP histogram Support vector machine (SVM) which is used to classify the normal and cancerous region. Leonardo de oliveira et al.[9],have been used k-means for segmentation and normal and cancerous tissue have been classified using support vector machine. Anuj Kumar et al.[10] proposed method, first detects the cancerous area, then segment the respective area. Average filter and thresholding method have been used for detection of cancerous region.S. Naresh et al. [11] employed Local Binary Pattern (CLBP) for extraction of texture feature and support vector machine (SVM) has been used for classification. Classification of six different types of breast cancer namely: CALC, CIRC, SPIC, MISC, ARCH, ASYM have been developed by Eanes Torres Peirrira et al.[12]. The proposed work is based on LBP feature and SVM classifier. An algorithm for mass detection has been developed by Arnau Oliver et al. [13], which is used for the reduction of false positives. Sun et al. [14] presented a new scheme based on multi view in the context of mammographic image analysis, which reduce the number of false classifications. The average of the classification results in two different views of the same image has been defined by this new method.

Manifold Learning method which focused for Density Segmentation in High Risk Mammograms has been proposed by Harry Strange et al. [15].QaisarAbbasa et al. [16], proposed dynamic contrast enhancement(DCE)which is used for pre-processing of mammogram. Dynamic contrast enhancement method has been used dynamic adaptive histogram equalization, Gaussian filtering and gamma correction techniques. Patricia B. Ribeiro et al. [17] introduced 'Optimum-Path Forest', a new technique for the classification of regions of interest in mammography images and a satisfactory performance has been obtained compared with other techniques. Combinations of random walks and active contour methods have been proposed by XinHAO et al.[18] for the mass segmentation on mammograms.

MATERIAL AND METHOD

In our study, MIAS (Mammographic Image Analysis Society) dataset which is standard and publicly available has been used for our experiment. Our proposed methodology has been evaluated by30 mammogram which belongs to two categories: benign and malignant. Out of 30 mammogram images 15 mammogram images have been diagnosed as normal and 15 as malignant. The size of each image is 1024X1024 pixels with 240 microns resolution.

A simple approach has been proposed for the detection of cancer region in mammogram. Our proposed methodology has been comprised of the three sequential steps as shown in figure 1.

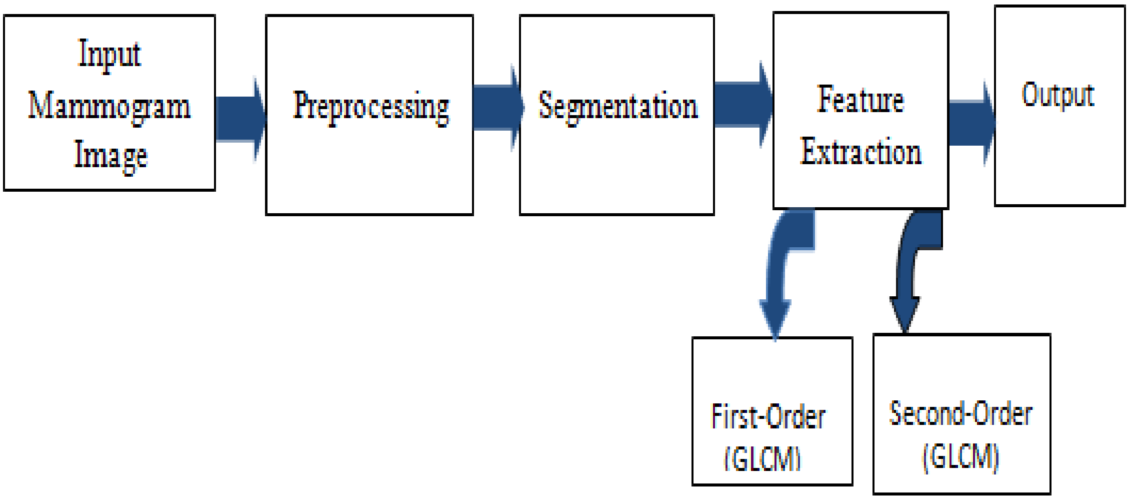


Figure 1: Block diagram of Proposed Method

PROPOSED METHOD

The three major steps are involved in the designing of the entire system such as (a) Preprocessing, (b) Segmentation, and (c) Feature Extraction. Detailed descriptions of steps for each module are as follows:

Preprocessing

Image preprocessing is the first and most important step as noise and unwanted distortions present in image, so to improve and enhance the image quality for human interpretation, preprocessing is required.

Procedures

- Convert the original mammogram image into gray scale image and image resizing (256 X 256) has been performed.
- Noise degrades the quality of image so noise removal is an essential step to restore and analyze the image to its original state.
- Homomorphic filter has been applied on image to enhance the contrast and compress the range of brightness. A hybrid approach has been proposed in homomorphic filtering process. The steps of hybrid approach are as follows:

  Step 1: Input noise free image.

  Step 2: Fourier Transform has been applied.

  Step 3: Log Transformation has been performed on the resultant image.

  Step 4: Inverse Fourier Transform has been done on the output image obtained from the log Transform.

  Step 5: Finally Exponential is applied to get the desired output.

- Top-hat and Bottom-hat transform have been used to smooth the borders and objects of the resultant image.
- Adaptive histogram equalization has been applied after combining the result of Top-hat and Bottom-hat transform will help to enhance the local contrast.

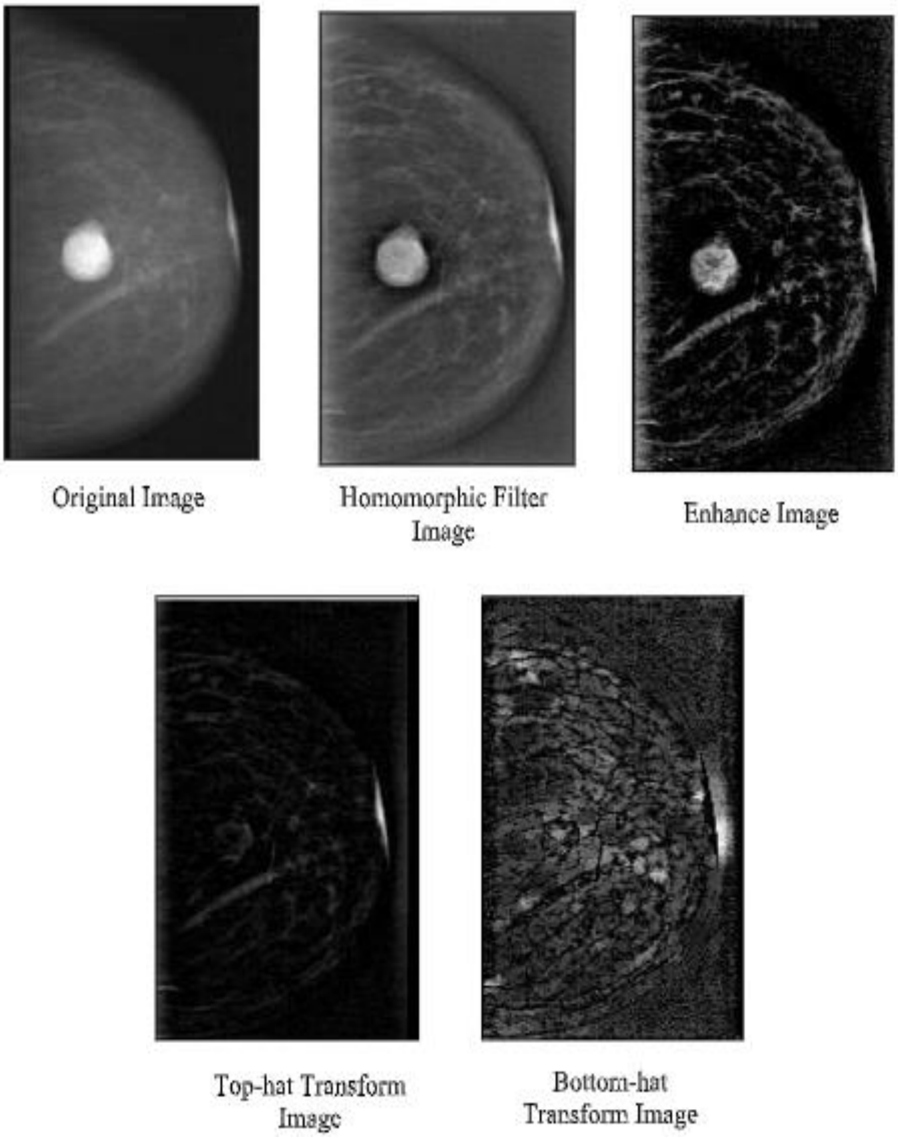
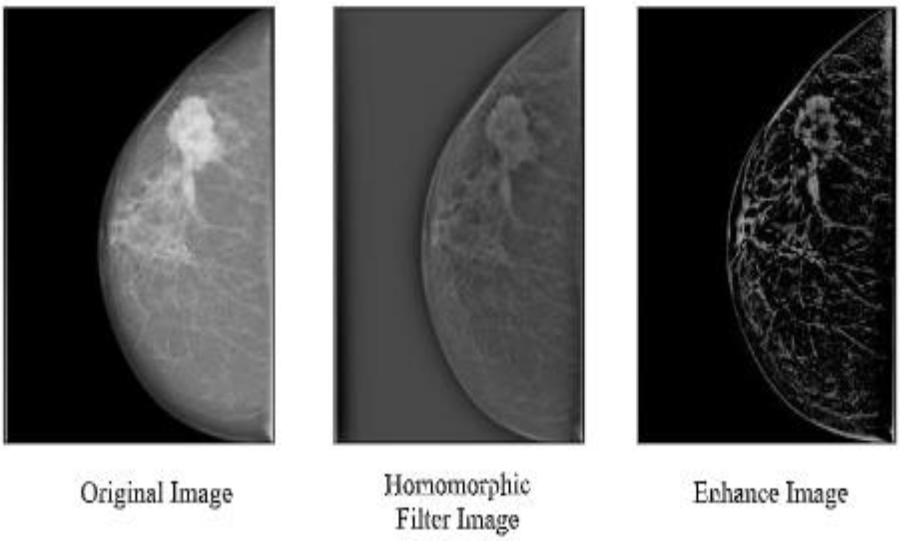


Figure 2: Preprocess Images of Benign Mass



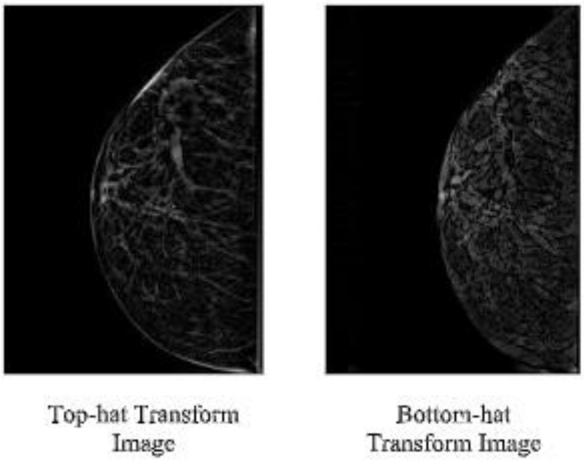


Figure 3: Preprocess Images of Malignant Mass

Segmentation

After performing the preprocessing step, ROI (Region of Interest) selection is the next major step. Background and pectoral muscles are not the part of breast so these unwanted regions must be removed to increase the performance as well as to focus on the region where the probability of cancer exist. The steps of segmentation are as follows:

Step1: Thresholding

Step2: Morphological operations and Edge detection

Step 3: K-Means Clustering.

Thresholding

Thresholding is the simplest method where pixels are partitioned depending on their intensity value. Each pixel is compared with a fixed constant value that is called the threshold. If the intensity value is higher than the threshold value then each pixel is replaced with white pixel and it is considered to be 'foreground', it is black pixel if the pixel value less than that constant value and it is considered as 'background' [19]. Thresholding has been used to eliminate unnecessary details and remove small objects.

Morphological Operation

Erosion and dilation have been used as morphological operations.

- Erosion: Eliminating the small details and highlighting the holes and gaps of different regions erosion has been used.
- Dilation: Dilation has been used to smooth the object boundary, close the holes and gaps and also expands the size of the object.

Edge Detection

The 'sobel' edge detection method has been used to find the edges. Two types of threshold values have been used to detect the strong and weak edges. Low edge sensitivity has been identified by high threshold value and low threshold value is used for high edge sensitivity.

K-Means

K-Means is the most popular partitioning method of clustering. This clustering technique classifies or groups different objects into k number of groups depending on attributes and features.

In this clustering algorithm for each k cluster k centre has to select randomly and k value is fixed in advance [20]. In a cluster, pixels having minimum distance attributes are grouped together. Euclidean distance has been used to calculate the distance between each data object and cluster centroid. Depending on the pixel of a cluster, a new centre has to calculate for each cluster. Pixels movement from one cluster to another cluster depends on the change of centroid in every step. This process will continue until no pixels will move from one cluster to another.

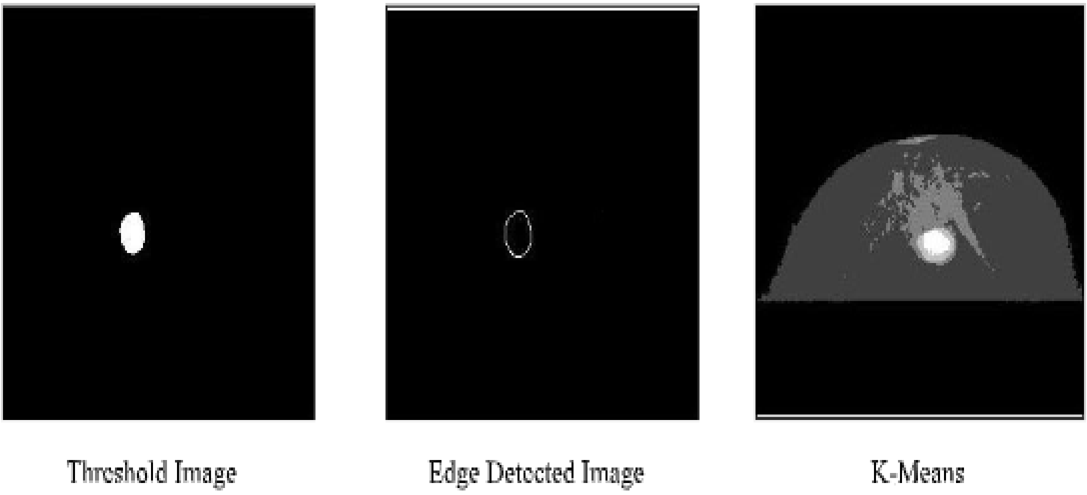


Figure 4: Segmented Images of Benign Mass

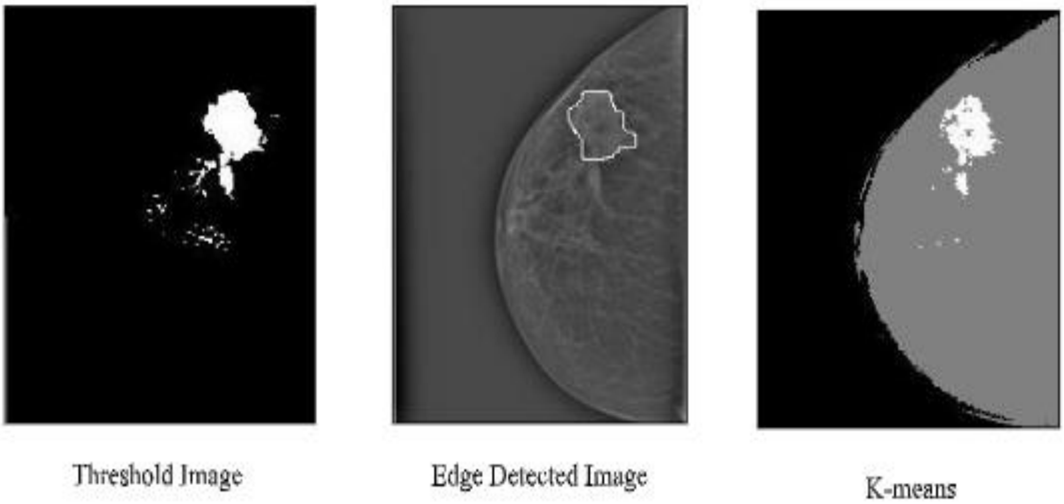


Figure 5: Segmented Images of Malignant Mass

Feature Extraction: Feature extraction is the most important decision making process for the detection of abnormalities of mammogram. Extracted features are very useful to differentiate the masses and normal breast tissue. First order and second order statistical based features have been extracted for benign and malignant tumor of mammogram images. The statistical features are defined by the following equations in the below table 1 and table 2.

Table 1 First Order Features

Moment	Definition	Formulae
Mean	The mean is the average value of all pixels in an image.	$\mu = \sum_{i=1}^{G-1} ip(i)$
Standard Deviation	It is the measurement of the average contrast	$\sigma = \sqrt{\sum_{i=0}^{G-1} (1-\mu)^2 (p(i))}$
Variance	It determines the intensity variation around the mean	$\sigma^2 = \sum_{i=1}^{G-1} (1-\mu)^2 (p(i))$
Kurtosis	Kurtosis is the fatness of the histogram	$\sigma^{-4} \sum_{i-1}^{G-1} (1-\mu)^4 (p(i))$

Table 2: GLCM (Gray Level Co-Occurrence Matrix) Features

Moment	Definition	Formulae
Contrast	It measures the sudden change of intensity value in image.	$\sum_{i,j} \lvert i-j \rvert^2 p(i.j)$
Correlation	It measures how the reference pixel correlates with its neighbor over an image.	$\sum_{i,j} \left(\frac{(i-\mu i)(j-\mu j) p(i,j)}{\sigma_i \sigma_j} \right)$
Energy	Energy is also called as angular second moment which defines the summation of squared elements in the GLCM.	$\sum_{i,j} pi,j^2$
Homogeneity	Homogeneity describes the distribution closeness of the element in the GLCM to GLCM diagonal.	$\sum_{i,j} \frac{p(i,j)}{i + \lvert i-j \rvert}$

RESULTS AND DISCUSSIONS

Table 3 and 4 represents the two different types of statistical based features first order and GLCM features (second order) for detection of abnormalities of mammogram. First order texture features such as, standard deviation, mean, variance, and kurtosis have been calculated from the image. Variance depends on the mean value and the mean value is low if the mass is benign, mean shows the brightness part in the image. Kurtosis shows a lower value for the malignant part. Standard deviation shows the lower value for malignant tumor and higher value for benign tumor.

Table 3: GLCM Features (First Order)

Type	Standard Deviation	Mean	Variance	Kutosis
Benign	0.2592	0.3290	0.0672	21.56
Malignant	0.5973	0.4658	0.8660	10.01

Table 4 represents the GLCM features. GLCM texture features like contrast, correlation, energy and homogeneity are calculated from the image. Contrast, correlation, energy and homogeneity are high for malignant tumor and low value for benign tumor.

Table 4: GLCM Features (Second Order)

Type	Contrast	Correlation	Energy	Homogeneity
Benign	0.5345	0.9432	0.0924	0.8344
Malignant	0.8660	0.9412	0.1051	0.7995

Signature

A distance has been calculated as a function of angle for benign and malignant mass from the centroid to boundary. Figure 6 shows the signature of boundary for benign and malignant mass.

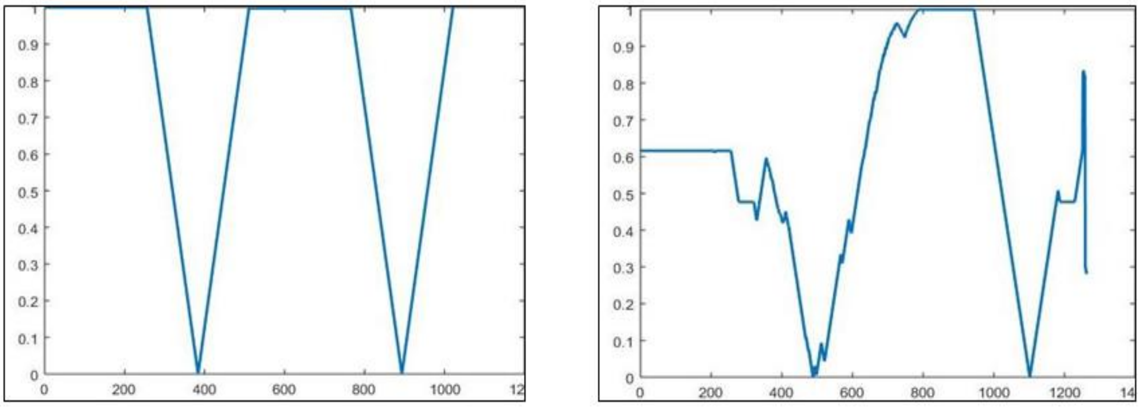


Figure 6: Boundary Signature for Benign and Malignant Mass

CONCLUSIONS AND FUTURE WORK

This paper enabled to develop an algorithm which exhibits a promising performance with capabilities of diagnosis of breast cancer. The aim of this paper is to enhance the accuracy of this proposed system for breast cancer detection. Statistical and GLCM both feature extraction methods have been used to increase the efficiency and accuracy of the proposed breast cancer detection system.

This system will be very helpful for the doctor for proper diagnose of breast cancer at an early stage and also capable to analyze the stage of cancer of a particular patient as that patient can get necessary, proper and fast treatment.

Future work will include the automated detection algorithm with classification using deep learning method which will reduces the computation time and increases the accuracy level more.

REFERENCES

1. *L. Hadjiiski, B. Sahiner, M. A. Helvie et al., "Breast masses: computer-aided diagnosis with serial mammograms," Radiology, vol. 240, no. 2, pp. 343–356, 2006.*
2. *Ranjana Giri, Subrat Sahu, K.L Purohit, J Rath &U Senapat, "Giant Congenital Nevus with Plexiform neurofibroma and Epitheloid Variant of Malignant Peripheral Nerve Sheath Tumor with Melanocytic Differentiation a Rare Case Report", International Journal of General Medicine and Pharmacy (IJGMP) ,. 4, Issue 2, pp, 81-86*

3. *S. W. Duffy, L. Tabar, and R. A. Smith, "The mammographic screening trials: commentary on the recent work by Olsen and Gotzsche," CA Cancer J Clin, vol. 52, no. 2, pp. 68–71, 2002.*

4. *Mohammad Omid Khairandish, Ruchika Gupta & Meenakshi Sharma, "A Hybrid Model of Faster R-CNN and SVM for Tumor Detection and Classification of MRI Brain Images", International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) ,Vol. 10, Issue 3,pp, 6863–6876*

5. *L. Tabar, B. Vitak, H. H. Chen, M. F. Yen, S. W. Duffy, and R. A. Smith, "Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality," Cancer, vol. 91, no. 9, pp. 1724–31, 2001.*

6. *P. Suresh Babu & B. Savitha, "Mri Brain Image Analysis for Tumor Detection Using Optimization Technique", International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) ,Vol. 3, Issue 3, pp, 23-28*

7. *A. N. Tosteson, D. G. Fryback, C. S. Hammond, L. G. Hanna, M. R. Grove, M. Brown, Q. Wang, K. Lindfors, and E. D. Pisano, "Consequences of false-positive screening mammograms," JAMA internal medicine, vol. 174, no. 6, pp. 954–961, 2014.*

8. *T. Balakumaran, I. L. A. Vennila, and C. G. Shankar, "Detection of microcalcification in mammograms using wavelet transform and fuzzy shell clustering," International Journal of ComputerScience and Information Technology, vol. 7, no. 1, pp. 121–125,2010.*

9. *Vrushali D. Dharmale & P. A. Tijare, "Cyst Detection in MRI Brain Image", IMPACT: International Journal of Research in Engineering & Technology (IMPACT: IJRET) , Vol. 2, Issue 2,pp, 129-138*

10. *R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," CA: a cancer journal for clinicians, vol. 65, no. 1, pp. 5–29, 2015.*

11. *R. Guzman-Cabrera, J.R. Guzaman-Supulveda, M. Torres-Cisneros, D.A. May-Arrioja, J. Ruiz-Pinales, O.G. Ibarra-Manzano, G. Avina-Cervantes, A. Donzalez Parada, "Digital image processing technique for breast cancer detection", Int J Thermophys(2013)34:1519-1531*

12. *Pavel Kral, Ladislav Lenc, "LBP features for breast cancer detection", ICIP2016, pp. 2643-2647.*

13. *Leonardo de Oliveira Martins, Geraldo Braz Junior, Aristifanes Correa Silva, Anselmo Cardoso de Paiva, and Marcelo Gattass, "Detection of masses in digital mammogram using K-means and support vector machine", Electronic Letters on Computer Vision and Image Analysis8(2):39-50, 2009.*

14. *Anuj Kumar Singh, Bhupendra Gupta, "A novel approach for breast cancer detection and segmentation in a mammogram", Eleventh International Multi-Conference on Information Processing-2015(IMCIP-2015), Procedia Computer Science 54(2015), pp. 676-682.*

15. *S. Naresh, S. Vani Kumari, "Breast Cancer Detection using Local Binary Patterns", International Journal Of Computer Applications (0975-8887), Volume 123- No.16, August 2015, pp. 6-9.*

16. *Eanes Torres Pereira, Sidney Pimentel Eleuterio, Joao Marques de Carvalho, "Local Binary Patterns Applied to Breast Cancer Classification in Mammographies", RITA Volume Numero, pp. 1-15.*

17. *Arnau Oliver, Xavier Llado, "False Positive Reduction in Mammographic Mass Detection Using Local Binary Patterns", MICCAI2007, Part I, LNCS 4791, pp.286-293, 2007.*

18. *L. Sun, L. Li, W. Xu, W. Liu, J. Zhang, and G. Shao, "A novel classification scheme for breast masses based on multi-view information fusion," in 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE), 2010.*

19. Harry Strange, Erika Denton, Minnie Kibiro, and Reyer Zwiggelaar, "Manifold Learning for Density Segmentation in High Risk Mammograms", IbPRIA 2013, pp. 245–252, 2013, Springer-Verlag Berlin Heidelberg 2013.

20. Qaisar Abbasa, M. Emre Celebic, Irene Fondon Garce, "Breast mass segmentation using region-based and edge-based methods in a 4-stage multiscale system", Biomedical Signal Processing and Control 8 (2013) 204–214, Elsevier.

21. Patricia B. Ribeiro, Roseli A.F. Romero, Patrícia R. Oliveira, Homero Schiabel, Luciana B. Verçosa, Automatic "segmentation of breast masses using enhanced ICA mixture model", Neurocomputing 120(2013) 61–71, Elsevier.

22. Xin HAO, Ye SHEN, Shun-ren XIA, "Automatic mass segmentation on mammograms combining random walks and active contour", J. Zhejiang Univ-Sci C (Comput& Electron) 2012 13(9):635-648, Springer.

23. Miss Hetal J. Vala, "A Review on Otsu Image Segmentation Algorithm", IJARCET, vol. 2, no. 2, pp. 387-389.

24. Shi Na, Liu Xumin, Guan Yong "Research on k-means Clustering Algorithm" Third International Symposium on Intelligent Information Technology and Security Informatics, IEEE 2010, pp. 63-67.